PromptMagician: Interactive Prompt Engineering for Text-to-Image Creation

Abstract

Generative text-to-image models have gained great popularity among the public for their powerful capability to generate high-quality images based on natural language prompts. However, developing effective prompts for desired images can be challenging due to the complexity and ambiguity of natural language. This research proposes PromptMagician, a visual analysis system that helps users explore the image results and refine the input prompts. The backbone of our system is a prompt recommendation model that takes user prompts as input, retrieves similar prompt-image pairs from DiffusionDB, and identifies special (important and relevant) prompt keywords. To facilitate interactive prompt refinement, PromptMagician introduces a multi-level visualization for the cross-modal embedding of the retrieved images and recommended keywords, and supports users in specifying multiple criteria for personalized exploration. Two usage scenarios, a user study, and expert interviews demonstrate the effectiveness and usability of our system, suggesting it facilitates prompt engineering and improves the creativity support of the generative text-to-image model.